Hyperbolic Geometry

James Stankewicz

September 28, 2015

1 Lecture I: Motivation and Basic definitions

It is an unfortunate truth that if you tell someone that you are studying mathematics, they may suddenly decide to ask if "someone didn't figure that out years ago," or about "the applications of your work." Especially with pure mathematics, it can be difficult to convey why you feel that your work is important, so I try to reframe the question in terms of the history of mathematics and science. Lately, the effects of even a breakthrough in pure mathematics have taken something like 100 years to see in the outside world. I'm fond of the example of GPS and Riemannian Geometry. The GPS that gets you around in a car was developed in the 1960s and 1970s by the US government and made available for unrestricted usage on the 2nd of May, 2000. Managing the GPS satellites would not be possible without the development of Einstein's general relativity, which celebrates its 100th birthday on November 18. But similarly to how GPS would not be possible without general relativity, the language of general relativity is that of Riemannian Geometry [4], whose basic structure was laid out by Riemann in a posthumous publication in 1867 [3]. So it may be quite a while before your great ideas make an impact on the world!

This course will focus on a precursor of even that idea. Before you can talk about a space with a curvature on it, you have to understand in a deep way what it means for a space's geometry to be anything other than the flat geometry of Euclid's *Elements*. This is something that I would call a *big idea* of modern mathematics, and amateur attempts to understand it have been known to end with literal Lovecraftian horror [2, Chapter 3]. Our approach will be the simple study of distances and angles on the hyperbolic upper half plane, which we will refer to as $\mathcal{H}$ and define as follows.

Let $\mathbf{C} = \mathbf{R}+i\mathbf{R}$ be the complex numbers where i is a fixed choice of a square root of -1. We may arrange $\mathbf{R} \subset \mathbf{C}$ as the horizontal axis in the real Euclidean plane and $i\mathbf{R}$ as the vertical axis. Recall formula for the length of an arc in the plane.

Definition 1. *We say that an* arc *in a topological space U is the image of a continuous map*

$$\gamma : [0,1] \to U.$$

We say that the length of an arc in the plane is

$$\int_\gamma \mathrm{d}s = \int_\gamma \sqrt{\mathrm{d}x^2 + \mathrm{d}y^2} = \int_0^1 \sqrt{\left(\frac{\mathrm{d}x(\gamma)}{\mathrm{d}t}\right)^2 + \left(\frac{\mathrm{d}y(\gamma)}{\mathrm{d}t}\right)^2} \mathrm{d}t.$$

It requires a proof to show that this definition conforms to our intuition, but that just comes down to the Pythagorean theorem. We say here that $\mathrm{d}s$ is the *metric* on the topological space $\mathbf{R}^2$. We say that the pair $(\mathbf{R}^2, \mathrm{d}s)$ is the *Euclidean plane.* One topic we will explore in this class is how much can change if we change the metric and how much we can do *even without a metric.*

Definition 2. *As a topological space, let $\mathcal{H} \subset \mathbf{R}^2 \cong \mathbf{C}$ be the subspace of complex numbers $x + iy$ with $y > 0$. We say that that* Poincare upper half-plane *is the space $\mathcal{H}$ together with the metric* $\dfrac{\sqrt{\mathrm{d}x^2 + \mathrm{d}y^2}}{y}$.

So what does this mean? Informally, this just means that things get closer together as we go upwards and further apart as we go downwards and closer to the line $y = 0$. More precisely, this means that what we think of as a straight line is going to change.

Definition 3. *An arc γ in a topological space U with a metric* $\mathrm{d}m$ *is called* geodesic *if it is the shortest path between two points $a = \gamma(0)$ and $b = \gamma(1)$. That is to say that for any other path γ' in U with $\gamma'(0) = a$ and $\gamma'(1) = b$ that the length $\ell(\gamma')$ satisfies $\ell(\gamma') \geq \ell(\gamma)$.*

Example 1. *In the Euclidean plane $(\mathbf{R}^2, \mathrm{d}s)$, the geodesics lie on "straight lines."*

Example 2. *On the sphere $\{(x, y, z) \in \mathbf{R}^3 : x^2 + y^2 + z^2 = 1\}$ with the usual Euclidean metric (restricted to the sphere!), the geodesics lie on* great circles.

In our next lecture, we will see that the geodesics in $\mathcal{H}$ lie on vertical lines and on semicircles whose center lies on the line $y = 0$. But for now, let's take a moment to wonder why we're calling this a metric and what the difference is to a metric space. It turns out we can define a metric using this "geodesic" or "least length" notion.

Definition 4. *If $z, w \in \mathcal{H}$, we let $\rho(z, w)$ be the infimum of the lengths of all arcs γ with $\gamma(0) = z$ and $\gamma(1) = w$. We say that $\rho(z, w)$ is the* hyperbolic distance *between z and w.*

Since this ρ is non-negative, symmetric, and satisfies the triangle inequality, it is a metric in the normal sense.

Definition 5. *We say that a continuous function $f : \mathcal{H} \to \mathcal{H}$ is an* isometry *if it preserves the hyperbolic distance. That is to say, for all $z, w \in \mathcal{H}$, $\rho(z, w) = \rho(f(z), f(w))$.*

2 Lecture 2: Linear Fractional Transformations

One of the great uses for topological spaces is the theory of group actions.

Definition 6. *We say that a group G acts on a set X if for all $g \in G$, $x \in X$, there is an element $g \cdot x \in X$ such that if $g, h \in G$, $x \in X$ then $g \cdot (h \cdot x) = (gh) \cdot x$ and the identity element of G fixes each element of X.*

If a group G acts nicely on a set X, that tells us something useful about the group G. Of course by nice, it's not enough to use the action $g \cdot x = x$ for all $g \in G$ and $x \in X$.

Last time we defined isometries, or transformations which preserve distance. The simplest example of these are *affine linear transformations* $\mathbf{R} \to \mathbf{R}$. Given $a \in \mathbf{R}^\times$ and $b \in \mathbf{R}$, consider the transformation sending $x \in \mathbf{R}$ to $ax+b$. Which of these preserve the usual distance? Of course the ones with $a = \pm 1$.

Consider now $a, b, c, d \in \mathbf{R}$ and $z \in \mathcal{H}$. We can try and make a distance-preserving map $\mathcal{H} \to \mathcal{H}$ sending z to $w = \dfrac{az+b}{cz+d}$. We run into an obvious problem if $ad = bc$ (and $c \neq 0$, something similar occurs if $a \neq 00$:

$$\frac{az+b}{cz+d} = \frac{acz+bc}{c(cz+d)} = \frac{a(cz+d)}{c(cz+d)} = \frac{a}{c}.$$

A map which collapses $\mathcal{H}$ to a point sends all distances to zero, and of course does not preserve distances. Therefore we assume $ad \neq bc$, or in terms that may be useful for us later, $ad - bc \in \mathbf{R}^\times$. If we express this transformation as a matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$, the group of these is called $\mathrm{GL}_2(\mathbf{R})$. We use this notation because these are not just affine linear transformations, but general linear transformations.

Another obvious thing to check is that this map sends $\mathcal{H}$ to $\mathcal{H}$, i.e. that $\Im(w) = \frac{w-\bar{w}}{2i} > 0$. We see clearly that

$$w = \frac{az+b}{cz+d} = \frac{ac|z|^2 + adz + bc\bar{z} + bd}{|cz+d|^2}.$$

It follows that

$$\Im(w) = \frac{(ad-bc)(z-\bar{z})}{2i|cz+d|^2} = (ad-bc)\frac{\Im(z)}{|cz+d|^2}.$$

Since $\Im(z) > 0$ and $|cz+d|^2 > 0$, the *linear fractional transformation* $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ sends $\mathcal{H}$ to $\mathcal{H}$ if and only if $ad - bc > 0$. The quantity $ad - bc$ is called the determinant of the matrix A and the set of matrices A with $\det(A) > 0$ forms a group, which we may denote as $\mathrm{GL}_2^+(\mathbf{R})$.

Each element of $\mathrm{GL}_2^+(\mathbf{R})$ defines a map from $\mathcal{H}$ to itself, which is continuous because polynomial maps and their quotients are continuous wherever they are defined. Finally, since $\mathrm{GL}_2^+(\mathbf{R})$ is a group, each element has an inverse, and so each element of $\mathrm{GL}_2^+(\mathbf{R})$ defines a homeomorphism from $\mathcal{H}$ to itself.

Alternately, we could just consider the matrices which have $ad - bc = 1$. These also form a group, which we call the *special linear group* or $\mathrm{SL}_2(\mathbf{R})$.

Lemma 1. *A linear fractional transformation of $\mathcal{H}$ given by $A \in \mathrm{SL}_2(\mathbf{R})$ is an isometry.*

Proof. Let $\gamma : [0,1] \to \mathcal{H}$ be a path and $A(\gamma)$ the image of that path under the transformation A. Let $w(t) = u(t) + iv(t) = A(\gamma)(t)$ so the length of $A(\gamma)$ is $\int_0^1 \frac{\sqrt{(\frac{\mathrm{d}u}{\mathrm{d}t})^2 + (\frac{\mathrm{d}v}{\mathrm{d}t})^2 \mathrm{d}t}}{v}$.

We note that $|\mathrm{d}w|^2 = \mathrm{d}w\overline{\mathrm{d}w} = (\mathrm{d}u + i\mathrm{d}v)(\mathrm{d}u - i\mathrm{d}v)$. Therefore $|\mathrm{d}w| = \sqrt{\mathrm{d}u^2 + \mathrm{d}v^2}$ so the length of γ is $\int_0^1 \frac{|\frac{\mathrm{d}w}{\mathrm{d}t}|\mathrm{d}t}{\Im(w)}$.

We compute via the quotient rule that

$$|\frac{\mathrm{d}w}{\mathrm{d}t}| = \left|\frac{\mathrm{d}w}{\mathrm{d}z}\frac{\mathrm{d}z}{\mathrm{d}t}\right| = \left|\frac{a(cz+d) - c(az+b)}{(cz+d)^2}\frac{\mathrm{d}z}{\mathrm{d}t}\right| = \left|\frac{(ad-bc)}{(cz+d)^2}\frac{\mathrm{d}z}{\mathrm{d}t}\right| = \frac{|\frac{\mathrm{d}z}{\mathrm{d}t}|}{|cz+d|^2}.$$

Recall then that $\Im(w) = \frac{\Im(z)}{|cz+d|^2}$, so

$$\int_0^1 \frac{|\frac{\mathrm{d}w}{\mathrm{d}t}|\mathrm{d}t}{\Im(w)} = \int_0^1 \frac{|\frac{\mathrm{d}z}{\mathrm{d}t}|\mathrm{d}t}{\Im(z)}.$$

Therefore the length of γ is the length of $A(\gamma)$. □

We therefore find that $\mathrm{SL}_2(\mathbf{R})$ acts on $\mathcal{H}$, not just continuously, but by isometries. This group action is nicer in other ways as well: It is *transitive* in that any point of $\mathcal{H}$ can be moved to any other point by an element of $\mathrm{SL}_2(\mathbf{R})$.

Lemma 2. *The action of $\mathrm{SL}_2(\mathbf{R})$ on $\mathcal{H}$ is transitive.*

Proof. Let $\alpha + \beta i, \alpha' + \beta' i \in \mathcal{H}$. We want to construct a matrix M taking $\alpha' + \beta' i$ to $\alpha + \beta i$. For simplicity, we assume that $(\alpha', \beta') = (0, 1)$ because if A is a matrix that takes i to $\alpha + \beta i$ and B takes i to $\alpha' + \beta' i$ then $M = B^{-1}A$ works.

Now if $\beta = 1$ then this is easy. Just take the matrix $\begin{pmatrix} \beta & \alpha \\ 0 & 1 \end{pmatrix}$. In general though, this matrix need not have determinant one because we only know that $\beta \in \mathbf{R}_{>0}$. We can correct this however - divide everything by $\sqrt{\beta}$! Then $A = \begin{pmatrix} \beta/\sqrt{\beta} & \alpha/\sqrt{\beta} \\ 0 & 1/\sqrt{\beta} \end{pmatrix}$ has determinant $\beta/(\sqrt{\beta}^2) = 1$ and takes i to $\frac{\frac{\beta i}{\sqrt{\beta}} + \frac{\alpha}{\sqrt{\beta}}}{\frac{1}{\sqrt{\beta}}} = \alpha + \beta i$. □

In fact, there is still a little more slack in this action as we will see in the next lecture when we look at geodesics.

3 Lecture 3: Linear fractional transformations and geodesics

We saw last time that we could move any point to any other point, e.g. to i. Let's see how much slack we have left after doing so. That is, once we get to i, which elements of $\mathrm{SL}_2(\mathbf{R})$ keep us there?

Lemma 3. *The elements of* $\mathrm{SL}_2(\mathbf{R})$ *which fix* $i \in \mathcal{H}$ *form the subgroup*

$$\mathrm{SO}_2(\mathbf{R}) = \left\{ \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} : \theta \in \mathbf{R} \right\}.$$

Proof. Suppose that $T = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbf{R})$. We may easiy compute that

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} i = \frac{ai+b}{ci+d} = \frac{(ad-bc)i + bd + ac}{c^2+c^2} = \frac{i}{c^2+d^2} + \frac{bd+ac}{c^2+d^2}.$$

Therefore the isometry induced by T fixes i if and only if $bd + ac = 0$ and $c^2 + d^2 = 1$. The latter equality together with the Pythagorean theorem tells us that there is some $\theta \in \mathbf{R}$ such that $c = \sin\theta$ and $d = \cos\theta$. The former equality now tells us that $a\sin\theta + b\cos\theta = 0$.

If $\cos\theta \neq 0$ we have $b = -a\tan\theta$. Since $1 = ad - bc = a(\tan\theta\sin\theta + \cos\theta)$, we have $\cos\theta = a(\sin^2\theta + \cos^2\theta)$. It follows that $a = \cos\theta$ and therefore $b = -\sin\theta$.

On the other hand if $d = \cos\theta = 0$ then $c = \sin\theta = \pm 1$. Therefore $a = \pm a\sin\theta = 0 = \cos\theta$ and $b = 1/c = \pm 1$. □

So what are we to make of this θ? Is there a way to see these matrices as a sort of rotation? Yes, in fact it may be helpful to think of i as being the center of a somewhat Dali-eqsue clock. If we pick a point $a + bi \neq i$, this is like choosing a minute hand, or rather a geodesic between i and $a + bi$ (whatever those look like!). Acting by an element of $\mathrm{SO}_2(\mathbf{R})$ rotates this minute hand. After doing so, we may assume that $a = 0$.

Exercise 1. *Find all the* θ *which move* $a+bi$ *to the imaginary axis. Check that if* $a^2+b^2 = 1$, *using* $\theta = \pi/4$ *does the job, while otherwise* $\frac{1}{2}\arctan(2a/(a^2+b^2-1))$ *does the job, moving* $a + bi$ *to the imaginary axis.*

We may also assume that $a + bi$ is moved to an element ir with $r > 1$. After all, if we take $\theta = 0$, we get the isometry $z \mapsto -1/z$ which sends ir to i/r.

Lemma 4. *The geodesics in* $\mathcal{H}$ *are all* $\mathrm{SL}_2(\mathbf{R})$*-equivalent to subsets of the imaginary axis* $\{ir : r \in \mathbf{R}_{>0}\}$.

Proof. Let γ be a path in $\mathcal{H}$ with beginning point z and ending point w. Apply an element M_1 of $\mathrm{SL}_2(\mathbf{R})$ to send z to i, and let $a + bi$ be the complex number w is sent to. Then apply an element M_2 of $\mathrm{SO}_2(\mathbf{R})$ so that overall $M = M_2M_1$

sends z to i and w to some ir for $r > 1$. Now let $M(\gamma)(t) = u(t) + iv(t)$ and consider that the length of γ is equal to the length of $M(\gamma)$, so

$$\int_0^1 \frac{\sqrt{\mathrm{d}u^2 + \mathrm{d}v^2}}{v} \geq \int_0^1 \frac{\mathrm{d}v}{v} = \ln(v(1)) - \ln(v(0)) = \ln(v(1)).$$

Therefore, any path between z and w has length at least $\ln(v(1))$. If we can make a path of that length between z and w, that will be a geodesic. Let $\sigma(t) = i((1-t) + tv(1))$, the vertical line going from i to $iv(1)$, which by the above has length $\ln(v(1))$. It follows that $M^{-1}(\sigma)$ is a geodesic arc between z and w. $\square$

Now what do the geodesics on $\mathcal{H}$ look like? There are the obvious vertical lines $\Re(z) = r$, which we can get by the matrix $\begin{pmatrix} 1 & r \\ 0 & 1 \end{pmatrix}$ or $\begin{pmatrix} r & -1 \\ 1 & 0 \end{pmatrix}$. The only other choices are half-circles centered on the real line: if $cd \neq 0$, then $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ sends the imaginary axis to the semicircle with edges $b/d, a/c$ on the real line.

Exercise 2. *Find the center, and radius of this half-circle. Find the highest point on this half-circle and the preimage of that point on the imaginary axis.*

But of course all of these things point to something that looks obvious in hindsight: we need to consider the limits of elements in $\mathcal{H}$. That is, we should throw in real numbers, and if we rotate around our clock, there's no good reason to throw out ∞. Now we should be careful: in general you simply can't treat ∞ like a number. You can however treat it like a point! And once we do that, you may as well throw in the lower half plane as well, and consider linear fractional transformations with arbitrary complex coefficients $a, b, c, d \in \mathbf{C}$ such that $ad - bc \neq 0$. Then on this completed space $\mathbf{C} \cup \{\infty\}$, the action of $\mathrm{GL}_2(\mathbf{C})$ is not just transitive, but 3-transitive in that any 3 points can be moved to any 3 other points (including the originals).

So for instance, if we wanted to move a point z_1 to zero, we would use the transformation $z \mapsto z - z_1$. If we wanted to move z_1 to zero and z_2 to ∞, we would use the transformation $z \mapsto \frac{z-z_1}{z-z_2}$. If we additionally want to map z_3 to 1, we use the transformation

$$z \mapsto \lambda(z_1, z_2, z_3, z) = \frac{z - z_1}{z - z_2} \Bigg/ \frac{z_3 - z_1}{z_3 - z_2}$$

This quantity $\lambda(z_1, z_2, z_3, z_4)$ is called the cross-ratio and it's a useful tool for coming up with linear transformations. For instance, if $\lambda = \lambda(z_1, z_2, z_3, z_4)$ then it is easy to compute that $\lambda = \lambda(0, \infty, 1, \lambda)$. Suppose now that $z, w \in \mathcal{H}$ and let C be the half-circle (or vertical line) containing both of them. Let $z^* \in \mathbf{R} \cup \infty$ be the endpoint closer to z and likewise w^* for w. Then we may interpret the distance between elements of $\mathcal{H}$ as

$$\rho(z, w) = \ln(\lambda(z^*, w^*, z, w)).$$

4 Lecture 4: The Unit Disc and the Iwasawa decomposition

We have used linear fractional transformations with complex coefficients to analyze the hyperbolic distance on the upper half plane. We now use them to give another model of hyperbolic geometry.

Definition 7. *The complex unit disc is the set* $\mathbb{D} = \{z \in \mathbf{C} : |z| < 1\}$*, together with the metric* $\frac{2|\mathrm{d}z|}{1-|z|^2}$.

Notice that just as distances get greater as you get closer in $\mathcal{H}$ to the real line, distances get greater as you get closer in $\mathbb{D}$ to the unit circle. This model is especially well-known from Escher's tesselations. Note the geodesics pictured are again the intersections of normal circles (or straight lines) with the unit disc.

Let's see how we can see that this is another copy of the upper half plane. Let $f(z) = i\frac{z-i}{z+i}$. This map sends 0 to $-i$, 1 to 1, ∞ to i, and i to 0. In short, it defines an invertible map $f : \mathcal{H} \to \mathbb{D}$. It's also easy to show that the metric on $\mathbb{D}$ is given obtained from the usual metric on the upper half plane. Simply compute $\frac{\mathrm{d}w}{\Im w}$ where $w = \frac{iz+1}{z+i}$.

That the unit disc can give the same model of hyperbolic geometry is an incarnation of a very important theorem in complex analysis called *Uniformization.* It essentially states that up to *conformal equivalence*, there are only 3 possible simply connected complex domains: the unit disc, the complex numbers, and the Riemann sphere. As you might guess, linear fractional transformations are conformal maps, and there are many criteria for being conformal, but in particular a map is conformal if angles are preserved.

Let's use this opportunity to think about angles in hyperbolic geometry. Perhaps you have heard that we can get triangles in hyperbolic geometry with an angle sum of less than 180 degrees. We will see this and much more, but first we should see that linear fractional transformations preserve angles. We will show this by proving a much bigger theorem that we've been slowly approaching.

Theorem 1 (The Iwasawa Decomposition)**.** *For any element* $A \in \mathrm{SL}_2(\mathbf{R})$*, there is an angle* $\theta = \theta(A)$*, a positive real number* $\lambda = \lambda(A)$*, and a real number* $x = x(A)$ *such that*

$$A = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} \lambda & 0 \\ 0 & 1/\lambda \end{pmatrix} \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}.$$

Proof. There will be a sketch here. For details, please see Keith Conrad's handout[1]. □

Corollary 1. *For any* $A \in \mathrm{SL}_2(\mathbf{R})$*, the associated linear fractional transformation of* $\mathcal{H}$ *preserves angles.*

[1] `http://www.math.uconn.edu/~kconrad/blurbs/grouptheory/SL(2,R).pdf`

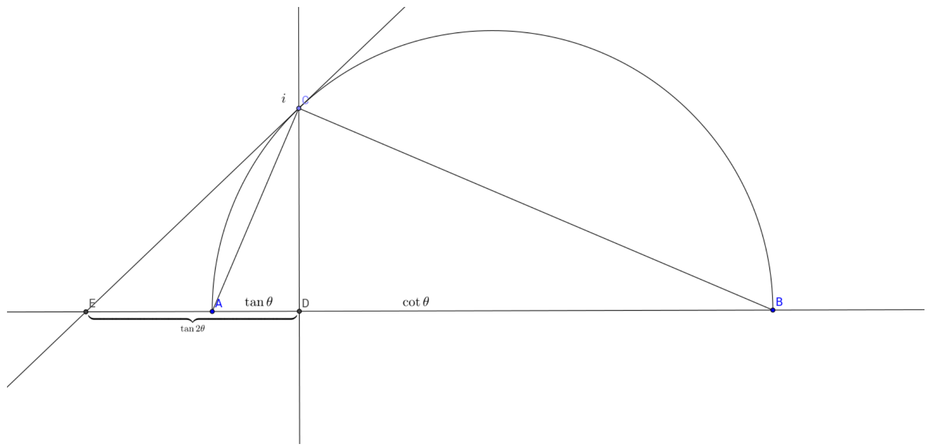

Figure 1: Exercise picture

Proof. Let's take two intersecting Euclidean straight lines and study what A does to the angle between the two. Note that matrices of the form $\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}$ simply move lines left and right. Moreover, any non-vertical line can be given in the form $y = ax + b$. Matrices of the form $\begin{pmatrix} \lambda & 0 \\ 0 & 1/\lambda \end{pmatrix}$ send z to $\lambda^2 z$. That is, x to $\lambda^2 x$ and y to $\lambda^2 y$. If $y = ax + b$ then after acting by this linear fractional transformation we have $y = ax + b/\lambda^2$. This is to say, the slopes of the two lines have not changed and so the angle between them has also not changed.

Now let P be the intersection point of two lines with angle α between them so we can study the resulting angle at $A(P)$. In fact, we can use a matrix M which is the product of matrices of the form $\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} \lambda & 0 \\ 0 & 1/\lambda \end{pmatrix}$ to move back to P. That is, $MA(P) = P$ and MA has the same effect on the angle α as A has. In fact, let N be a matrix formed in the same way as M, but such that $N(P) = i$. Then the angle α at P is just a translate under N^{-1} of the same angle at i. We know that $NMAN^{-1}$ fixes i (so lies in $\mathrm{SO}_2(\mathbf{R})$) and has the same effect on the angle α at i as A does at P. So we're done if we interpret the lines forming the angle α as tangent lines to geodesic circles and we know the effect of $\mathrm{SO}_2(\mathbf{R})$.

In fact, since the geodesics going through i are all equivalent under $\mathrm{SO}_2(\mathbf{R})$, it suffices to see what the angle is between the imaginary axis, and the tangent line at i to the image under the imaginary axis under $k_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$. In fact, if we can prove the following exercise, we can show the angle is 2θ. $\square$

Exercise 3. *Show that the lengths in the following picture are accurate if the semicircle is the image of the positive imaginary axis under* k_θ*.*

See also Keith Conrad's handout for a proof that $\mathrm{SL}_2(\mathbf{R})$ is homeomorphic to a solid torus and $\mathcal{H} \cong \mathrm{SL}_2(\mathbf{R})/\mathrm{SO}_2(\mathbf{R})$.

5 Lecture 5: Hyperbolic trigonometry

We will now study the properties of hyperbolic triangles. In the same way that a regular triangle is formed by taking the straight lines between 3 points in the plane, hyperbolic triangles are the same, *mutatis mutandis*. But here, we have an easily-defined boundary, and we can extend our definition to this boundary as follows.

Definition 8. *A* hyperbolic triangle *is formed by 3 vertices, which are distinct points in $\mathcal{H}$, together with 3 edges, which are all the hyperbolic geodesics arcs between these 3 points. An* ideal hyperbolic triangle *is the same, but the vertices can also be ∞ or any point in $\mathbf{R}$.*

The angles between two geodesics are given simply as the angles between the tangent lines of these geodesics, and there is one more key ingredient to studying hyperbolic trigonometry.

Lemma 5. *The area of a region $R \subset \mathcal{H}$ is $\iint_R \frac{\mathrm{d}x\mathrm{d}y}{y^2}$.*

Proof. This is an exercise. Explain why it suffices to check this on Euclidean rectangles $[a, b] \times [c, d]$ and then do so by usual integration. □

Example 3. *Let's see how far we can push this idea of an ideal triangle. Can we have a triangle with all points on the boundary? Of course! Let the 3 vertices be ∞, 1, and -1, so the edges are the lines $x = 1$, $x = -1$, and the top half of the unit circle.*

What is the area of this triangle? Well of course it is

$$\int_{-1}^{1} \mathrm{d}x \int_{\sqrt{1-x^2}}^{\infty} \frac{\mathrm{d}y}{y^2} = \int_{-1}^{1} \frac{\mathrm{d}x}{\sqrt{1-x^2}} = \int_{\pi}^{0} \frac{\mathrm{d}(\cos\theta)}{\sin\theta} = \int_{0}^{\pi} \mathrm{d}\theta = \pi.$$

More generally, if the edges of an ideal triangle are given by the unit circle, $x = a$, and $x = b$ for $-1 \leq a < b \leq 1$ then the area is $\arccos(a) - \arccos(b)$ by the above calculation.

Let the interior angle at $b + i\sqrt{1-b^2}$ be β and the interior angle at $a + i\sqrt{1-a^2}$ be α. We check with a little bit of standard Euclidean trigonometry that $\arccos(b) = \beta$ and $arccos(a) = \pi - \alpha$. Therefore in terms of interior angles, the area of our triangle with a vertex at ∞ is $\pi - \alpha - \beta$. In fact, it's easy to see that the angle at ∞ is zero as well - act by the linear fractional transformation $-1/z$ and recall that linear fractional transformations preserve angles. So any of these ideal triangles with interior angles $\{\alpha, \beta, \gamma\}$ has area $\pi - \alpha - \beta - \gamma$.

If this held true for all hyperbolic triangles, that would very precisely say how any hyperbolic triangle has angle sum less than π radians.

Let's think of how we can reduce all triangles to the above example. Since the hyperbolic metric is preserved by the action of $\mathrm{SL}_2(\mathbf{R})$, the hyperbolic area ought to be as well. More precisely, we know from multivariable calculus that if $f(x,y) = (u,v)$ then $\mathrm{d}u\mathrm{d}u = \left(\frac{\partial u}{\partial x}\frac{\partial v}{\partial y} - \frac{\partial u}{\partial y}\frac{\partial v}{\partial x}\right)\mathrm{d}x\mathrm{d}y$.

If $(u,v) = f(x,y) = f(x+iy) = u+iv$ has a complex derivative at a point z_0 in $\mathcal{H}$ then

$$f'(z_0) = \frac{\partial u}{\partial x}(z_0) + i\frac{\partial v}{\partial x}(z_0) = \frac{\partial v}{\partial y}(z_0) - i\frac{\partial u}{\partial y}(z_0).$$

Exercise 4. *Prove this. As a hint, let $z_0 = x_0 + iy_0$. Take the real two-variable directional derivatives along the horizontal line $y = y_0$ and along the vertical $x = x_0$. What should be true there and why?*

If our function f is a linear fractional transformation, then it has a complex derivative. In fact, if it's given by $T = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbf{R})$, then the derivative is $f'(z) = \frac{1}{(cz+d)^2}$ and so the Jacobian factor is $|f'(z)|^2 = \frac{1}{|cz+d|^4}$. Therefore,

$$\iint_{T(R)} \frac{\mathrm{d}u\mathrm{d}v}{v^2} = \iint_R \frac{|\, cz+d \,|^4}{y^2}\frac{\mathrm{d}x\mathrm{d}y}{|\, cz+d \,|^4} = \iint_R \frac{\mathrm{d}x\mathrm{d}y}{y^2}.$$

Now we can prove the following.

Theorem 2 (Gauss-Bonnet)**.** *The area of a hyperbolic triangle or ideal hyperbolic triangle with interior angles $\{\alpha,\beta,\gamma\}$ is $\pi - \alpha - \beta - \gamma$.*

Proof. First, we assume that we have at least one vertex in $\mathbf{R} \cup \infty$. In fact, we can assume this vertex is at ∞. If not, it is at some $x \in \mathbf{R}$, and the matrix $\begin{pmatrix} 0 & -1 \\ 1 & -x \end{pmatrix}$ moves x to ∞. Two sides will be vertical lines above the remaining vertices which lie in either $\mathcal{H}$ or $\mathbf{R}$. The remaining vertices lie on the closure of a geodesic arc, given as a half-circle centered on the real line. Let $y \in \mathbf{R}$ be the center and r^2 be the radius. The matrix $\begin{pmatrix} 1 & -y \\ 0 & 1 \end{pmatrix}$ centers this geodesic at 0 and the matrix $\begin{pmatrix} 1/r & 0 \\ 0 & r \end{pmatrix}$ leaves us with the unit circle. We are now in the situation of Example 3, where we know the theorem holds.

Now assume we have a true hyperbolic triangle with vertices A, B, and $C \in \mathcal{H}$ with associated angles α, β, and γ. Let $D \in \mathbf{R}$ be the limit along the geodesic arc between A and B so that the angle CBD is $\pi - \beta$. Let θ be the interior angle at C so that the triangle with vertices A, C, D has interior angles $\alpha, \gamma + \theta$, and zero. This is an ideal triangle and so we know the area is $\pi - \alpha - \gamma - \theta$. The smaller ideal triangle with vertices C, B, and D has interior angles $\theta, \pi - \beta$, and zero and so has area $\pi - (\pi - \beta) - \theta = \beta - \theta$. Taking the difference of these two areas gives us our result. $\square$

The full Gauss-Bonnet theorem is a much more general theorem in Riemannian geometry. Another important special case is Girard's theorem about triangles on spheres. On a sphere every triangle has an angle sum of more than π radians. Well on the sphere of radius 1, a triangle with interior angles α, β, and γ has area $\alpha + \beta + \gamma - \pi$. Here we can perhaps see a bit of what it means for $\mathcal{H}$ to be *negatively curved* and the sphere to be *positively curved*.

6 Lecture 6: The isometry group of $\mathcal{H}$

Way back in Lecture 2, we stated that we can tell a lot about a space X by considering a group G with a nice action on X. The action of $\mathrm{SL}_2(\mathbf{R})$ on $\mathcal{H}$ is nice for many reasons. It is transitive, it is continuous, it preserves areas, angles, and distances. On account of the last of these, the action gives a homomorphism $\mathrm{SL}_2(\mathbf{R}) \to \mathrm{Isom}(\mathcal{H})$.

There are other properties we might like from groups acting on a set that we do not have. For one, the action is not *faithful* in that there are different elements of $\mathrm{SL}_2(\mathbf{R})$ which act in the exact same way on $\mathcal{H}$. Since $\mathrm{SL}_2(\mathbf{R})$ is a group, this equivalently means that there are elements besides $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ which act as the identity isometry on $\mathcal{H}$ or that our homomorphism has a nontrivial kernel. It's pretty small though, so our action is not too far from being faithful.

Lemma 6. *The kernel of the homomorphism* $\mathrm{SL}_2(\mathbf{R}) \to \mathrm{Isom}(\mathcal{H})$ *is the subgroup* $\left\{ \begin{pmatrix} \pm 1 & 0 \\ 0 & \pm 1 \end{pmatrix} \right\}$ *which we may refer to as* $\{\pm 1\}$.

Proof. If $T \in \mathrm{SL}_2(\mathbf{R})$ acts as the identity, it must fix i and thus lie in $\mathrm{SO}_2(\mathbf{R})$. An element $k_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \in \mathrm{SO}_2(\mathbf{R})$ acts on the imaginary axis, rotating it by an angle of 2θ to another geodesic. Therefore the angle θ is an integer multiple of π and the statement is proved. $\square$

It now follows if $T \in \mathrm{SL}_2(\mathbf{R})$ that $-T$ transforms $\mathcal{H}$ in exactly the same way and every other element of $\mathrm{SL}_2(\mathbf{R})$ acts differently. So if we want to consider elements of $\mathrm{SL}_2(\mathbf{R})$ in terms of their action on $\mathcal{H}$, we should not consider single matrices A but pairs of matrices $\pm T$.

Put differently, we should use what is sometimes referred to as the *first isomorphism theorem*. That is, we have a homomorphism $\mathrm{SL}_2(\mathbf{R}) \to \mathrm{Isom}(\mathcal{H})$ with kernel $\{\pm 1\}$. This gives us a *one-to-one* or *injective* homomorphism $\mathrm{SL}_2(\mathbf{R})/\{\pm 1\} \hookrightarrow \mathrm{Isom}(\mathcal{H})$. We refer to the group $\mathrm{SL}_2(\mathbf{R})/\{\pm 1\}$ as the *projective special linear group* or $\mathrm{PSL}_2(\mathbf{R})$. Now $PSL_2(\mathbf{R})$ acts *faithfully* on $\mathcal{H}$ in that for all $\pm T \in \mathrm{PSL}_2(\mathbf{R})$, there is some $z \in \mathcal{H}$ such that $\pm Tz \neq z$.

Now how do we characterize elements of $\mathrm{PSL}_2(\mathbf{R})$? The Iwasawa decomposition points us in the right direction. It states that we have groups $K \cong \mathrm{SO}_2(\mathbf{R})$, $A \cong \mathbf{R}_{>0}$, and $N \cong \mathbf{R}$ such that $\mathrm{SL}_2(\mathbf{R}) = KAN$. Recall that an element $\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \in N$ always has trace 2 and has no fixed points in $\mathcal{H}$ unless $x = 0$. An element $\begin{pmatrix} r & 0 \\ 0 & 1/r \end{pmatrix}$ also has no fixed points in $\mathcal{H}$ unless $r = 1$, but the trace is $r+1/r \geq 2$. Of course an element $k_\theta \in K$ fixes i and no other elements unless θ is an integer multiple of π. This is the only subgroup which contains -1. The trace of k_θ is $2\cos\theta \leq 2$, but in fact we have $|2\cos\theta| \leq 2$. Now we have hit on something that is defined on $\mathrm{PSL}_2(\mathbf{R})$ - if $T \in \mathrm{SL}_2(\mathbf{R})$ has trace t, then $-T$ has trace $-t$ so the absolute value of the trace is a continuous function $\mathrm{PSL}_2(\mathbf{R}) \to \mathbf{R}_{\geq 0}$.

Definition 9. *If $T \in \mathrm{PSL}_2(\mathbf{R})$, we say that T is* elliptic *if $|\ \mathrm{tr}T\ | \leq 2$. We say that T is* parabolic *if the trace is equal to 2, and* hyperbolic *if it is ≥ 2.*

This quantity is especially nice because it will only see the conjugacy class of an element in $\mathrm{PSL}_2(\mathbf{R})$. That is, if $S, T \in \mathrm{PSL}_2(\mathbf{R})$ then $|\mathrm{tr}(STS^{-1})| = |\mathrm{tr}(T)|$. If for instance, if S sends i to $z \in \mathcal{H}$ then $S^{-1}(\mathrm{SO}_2(\mathbf{R})/\pm 1)S$ is the stabilizer subgroup of z.

Let's return to the issue of fixed points. Let $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbf{R})$ and $z \in \mathbf{C}$. We have z fixed if and only if it is a root of $cz^2 + (d-a)z - b = 0$. The discriminant of this polynomial is $\Delta = (a+d)^2 - 4$ and so there are two real roots if $\Delta > 0$, two imaginary roots if $\Delta < 0$ and a single real root if $\Delta = 0$. The upshot here is that unless T is elliptic, T has no fixed points in $\mathcal{H}$. And if all elements of a subgroup $\Gamma \leq \mathrm{PSL}_2(\mathbf{R})$ have no fixed points in $\mathcal{H}$ then the quotient space $\Gamma \setminus \mathcal{H}$ is often very nice (we can be more specific later).

Example 4. *Consider the parabolic subgroup $\mathbf{Z} \cong \Gamma \leq \mathrm{PSL}_2(\mathbf{R})$ generated by $\pm\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. The quotient map $\mathcal{H} \to \Gamma \setminus \mathcal{H}$ can be expressed in a surprisingly easy manner. Namely, if we can find a space X and a continuous function $f : \mathcal{H} \to X$ such that $f(z) = f(z+1)$ then we have a continuous map $\Gamma \setminus \mathcal{H} \to X$. If we can invert f enough to land in $\Gamma \setminus \mathcal{H}$, we can identify f with the quotient map. In this case $X = \mathbb{D}^\bullet = \mathbb{D} - \{0\}$ and $f(z) = e^{2\pi i z}$. Inverting f just enough to be defined on $\Gamma \setminus \mathcal{H}$ is just taking a branch cut of the logarithm, i.e. restricting the complex logarithm to have an argument in $[0, 2\pi)$.*

In the above example, we note that we take the quotient by a group isomorphic to $\mathbf{Z}$ and get a space whose fundamental group is $\mathbf{Z}$. This is an instance of something you'll see in more detail if you take algebraic topology. Namely, this group action is *wandering*, or satisfies what is often called a "covering space" condition. A group G acting on a path-connected space X by continuous maps satisfies this condition if for all $x \in X$ there is a small enough neighborhood U of x such that if $g \in G$, $gU \cap U \neq \emptyset$ implies $g = 1$. This is to say that G pushes open sets away from each other. Clearly the above action of $\mathbf{Z}$ on $\mathcal{H}$ satisfies this condition. If this condition is satisfied and X is simply connected, like $\mathcal{H}$, then $\pi_1(G \setminus X) \cong G$.

We should note that simply avoiding elliptic elements is not a panacea. For instance, the parabolic subgroup $N \cap \mathrm{SL}_2(\mathbf{Q})/\pm 1$ does not have a Hausdorff quotient. Moreover, leaving out elliptic elements prevents us from looking at some very interesting groups!

Example 5. *The group* $\mathrm{SL}_2(\mathbf{Z}) \leq \mathrm{SL}_2(\mathbf{R})$ *acts on* $\mathcal{H}$*, but* $\mathrm{SL}_2(\mathbf{Z}) \cap \mathrm{SO}_2(\mathbf{Z}) = \left\langle \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \right\rangle \cong \mathbf{Z}/4$*. Therefore if we quotient out by* ± 1*,* $\mathrm{PSL}_2(\mathbf{Z})$ *contains a* $\mathbf{Z}/2$ *stabilizing* i*. Can you find the other elliptic elements of* $\mathrm{PSL}_2(\mathbf{Z})$*?*

7 Lecture 7: Proper discontinuity

Finally we arrive at an important definiton, which will guarantee good topological properties for our quotient.

Definition 10. *We say that a subgroup* $\Gamma \leq \mathrm{PSL}_2(\mathbf{R})$ *is* Fuchsian *if the subspace topology on* Γ *is discrete.*

It's easy to see why any point $z \in \mathcal{H}$ has a finite stabilizer subgroup in Γ. The stabilizer of z in $\mathrm{PSL}_2(\mathbf{R})$ is conjugate to $\mathrm{SO}_2(\mathbf{R})/\pm 1 \cong \mathbf{R}/\pi\mathbf{Z}$, which is compact. Therefore Γ intersected with any of these stabilizers is a discrete subgroup of a compact space, and is therefore finite. Of course just having finite stabilizers is not enough to guarantee a good quotient, but it turns out that elements of Fuchsian groups move open sets away from each other unless they fix a point- they act *properly discontinuously.*

Definition 11. *We say that an action of a group* G *on a topological space* X *is* properly discontinuous *if for all compact sets* $K_1, K_2 \in X$*, we have* $gK_1 \cap K_2 = \emptyset$ *for all but finitely many* $g \in G$*.*

Theorem 3. *A group* G *acting properly discontinuously on a locally compact Hausdorff space* X *has a Hausdorff quotient* $G \setminus X$*.*

Proof. Suppose that $x_1, x_2 \in X$ such that $x_2 \notin Gx_1$, so the orbits Gx_1 and Gx_2 are distinct in $G \setminus X$. Since X is locally compact and Hausdorff, there are non-overlapping open sets V_i and compact sets K_i such that $x_i \in V_i \subset K_i$. Let $\{g_1, \ldots, g_n\}$ be the elements of G such that $g_j K_1 \cap K_2 \neq \emptyset$.

Without loss of generality, each of these g do not fix x, for if $gx = x$ then we can replace K_1 by the compact set $K_1' = K_1 \cap g^{-1}K_1$ and see that

$$gK_1' \cap K_2 = (gK_1 \cap K_1) \cap K_2 = gK_1 \cap (K_1 \cap K_2) = gK_1 \cap \emptyset = \emptyset.$$

Furthermore, we may assume that for each $g \in \{g_1, \ldots, g_n\}$ we have $x_2 \notin gK_1 \cap K_2$, for if not, replace K_1 with $K_1 - g^{-1}V_2$. Finally we see that if we let $V_2' = V_2 - \bigcup_j g_j K_1$, then for all $g \in G$, gV_1 is disjoint from V_2'. The quotient map $q : X \to G \setminus X$ was constructed so that if $U \subset X$ is open, then $q(U)$ is open. Therefore $q(V_1)$ and $q(V_2')$ are disjoint neighborhoods of Gx_1 and Gx_2. $\square$

Theorem 4. *Fuchsian groups act properly discontinuously on* $\mathcal{H}$*.*

Proof. Fix K_1, K_2 compact subsets of $\mathcal{H}$. It suffices to show that the set of all $\gamma \in \mathrm{SL}_2(\mathbf{R})$ such that $\gamma K_1 \cap K_2 \neq \emptyset$ is a compact set K. If so, we recall that the quotient map $q : \mathrm{SL}_2(\mathbf{R}) \to \mathrm{PSL}_2(\mathbf{R})$ is continuous and so $q(K)$ is also compact. It follows then that $\Gamma \cap q(K)$ is a discrete subset of a compact space and is therefore finite. If we let K_1 and K_2 range over all pairs of compact subsets of $\mathcal{H}$, we see that the action is therefore properly discontinuous.

So now we need to show that K is compact. To this end, recall that $\mathcal{H} \cong \mathrm{SL}_2(\mathbf{R})/\mathrm{SO}_2(\mathbf{R})$ and let $K_i' = K_i \times \mathrm{SO}_2(\mathbf{R})$, a compact subspace of $\mathrm{SL}_2(\mathbf{R})$. We do this to make the machinery of topological groups do the work for us.

We can therefore say that $\gamma \in K$ if and only if there exist $k_1 \in K_1$, $k_2 \in K_2$, $s_1, s_2 \in \mathrm{SO}_2(\mathbf{R})$ such that $\gamma k_1 s_1 = k_2 s_2$, or $\gamma = k_2 s_2 (k_1 s_1)^{-1}$. We see therefore that $K = K_2'(K_1')^{-1}$, and it is now easy to see that this is compact. Namely, since inversion and multiplication are continuous maps, they send compact sets to compact sets. $\square$

Exercise 5. *Show that if Γ is a torsion-free Fuchsian group, for all $z \in \mathcal{H}$ there is a neighborhood U of z such that if $g \in \Gamma$ and $gU \cap U \neq \emptyset$ then g is the identity.*

We see that the quotient of the upper half-plane by a Fuchsian group Γ gives us a nice Hausdorff quotient $\Gamma \setminus \mathcal{H}$. How do we understand that quotient? One way is by studying the way that $\mathcal{H}$ is *tessalated* by the action of Γ.

What we mean when we say that a Fuchsian group Γ tessalates $\mathcal{H}$ is that there is a closed, contractible subset $\mathcal{F}$ of $\mathcal{H}$ such that

1. $\bigcup_{\gamma \in \Gamma} \gamma \mathcal{F} = \mathcal{H}$,
2. $\mathcal{F}$ has nonempty interior, and
3. $\gamma \mathcal{F}^o \cap \gamma' \mathcal{F}^o = \emptyset$ unless $\gamma = \gamma'$.

We will frequently be interested in the case where $\mathcal{F}$ is a hyperbolic polygon, i.e. a union of finitely many possibly ideal hyperbolic triangles. If we have such a set, which we call a fundamental domain, then we can form the quotient $\Gamma \setminus \mathcal{H}$ by simply gluing together the appropriate sides of $\mathcal{F}$.

Example 6. *Take the ideal triangle with vertices $1, -1,$ and ∞. The interior is a fundamental domain for the group generated by $\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$.*

Really though, the best way to understand the quotient is with numbers, and the best way to obtain those numbers is by integrating along paths. So far the only thing that we've integrated is the metric $\sqrt{\mathrm{d}x^2 + \mathrm{d}y^2}/y = \mid \mathrm{d}z \mid /y$.

While the metric descends from $\mathcal{H}$ to $\Gamma \setminus \mathcal{H}$ if Γ is torsion-free, this is no longer the case if Γ has torsion. Especially, if $z_0 \in \mathcal{H}$ is fixed by $g \in \Gamma$ with $g^n = 1$ then locally at z_0, the quotient map looks like $z \mapsto z^n$. That is to say, we have a pinch point at z_0 where the metric breaks down.

What will always descend down however are holomorphic differential k-forms that are invariant under the action of Γ. We will say that a function on a region $f(z) = u(z) + iv(z)$ is holomorphic if $\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}$ and $\frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}$. That is, if every directional derivative of f is the same.

Definition 12. *A holomorphic differential n-form on a region R is something that can be integrated n-times, i.e. something of the form* $f(z)(\mathrm{d}z)^n$ *where* $f(z)$ *is holomorphic on R.*

Recall now that $\mathrm{d}\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix} z\right) = \frac{\mathrm{d}z}{(cz+d)^2}$. To ensure the whole differential $f(z)\mathrm{d}z$ is invariant under the action of Γ, we need $f\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix} z\right) = (cz+d)^{2n} f(z)$ for all elements of Γ.

8 Lecture 8: $\mathrm{SL}_2(\mathbf{Z})$ and modular forms

We want to consider differential forms on quotients of $\mathcal{H}$ by Fuchsian groups. These naturally lead to a very important class of functions.

Definition 13. *Let* f *be a holomorphic function on* $\mathcal{H}$. *We say that* f *is a weak modular form of weight* k *if for all* $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$, $f\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix} z\right) = (cz+d)^k f(z)$. *If additionally* $\lim_{\Im(z)\to\infty} f(z)$ *exists, we say it is a* modular form *of weight* k *for* Γ.

We may relax the limit requirement if $\Gamma \setminus \mathcal{H}$ is compact. It is possible to consider modular forms of odd weight, and they are very interesting depending on the group, but for the group $\mathrm{SL}_2(\mathbf{Z})$ or its image $\mathrm{PSL}_2(\mathbf{Z})$ in $\mathrm{PSL}_2(\mathbf{R})$ they are very simple. For instance if we take the matrix $S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ and k is odd, $f(z) = f(S^2 z) = (-1)^k f(z) = -f(z)$ and so f is constantly zero. However, if we take k even, we can get some very nice properties. For instance, if we take $T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ we see that $f(z) = f(z+1)$ so we have a Fourier expansion for f. Namely if we change variables to $q = e^{2\pi i z}$, we can write f as $f(q) = \sum_{n\geq 0} a_n q^n$. It turns out that any integer matrix with determinant one can be written in terms of these two.

Theorem 5. *The group* $\mathrm{SL}_2(\mathbf{Z})$ *is generated by* S *and* T.

Definition 14. *The* modular fundamental domain *(which we will see is the fundamental domain for* $\mathrm{SL}_2(\mathbf{Z})$*) is*

$$\mathcal{F} = \{x + iy \in \mathcal{H} : -1/2 \leq x \leq 1/2, x^2 + y^2 \geq 1\}.$$

Proof. Let $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in G = \langle S, T\rangle \leq \mathrm{SL}_2(\mathbf{Z})$. As we know, if $z = x + iy \in \mathcal{H}$ then

$$\Im(gz) = \frac{\Im(z)}{\mid cz + d \mid^2} = \frac{y}{(cx+d)^2 + (cy)^2}.$$

Since $y > 0$, there are only finitely many c making $|cz+d|^2 < B$ for any given real number B. Similarly if we fix one of those finitely many c, there are only finitely many d making $|cz+d|^2 < B$. For instance if we take $B=1$ there are only finitely many pairs (c,d) such that $\Im(gz) > Im(z)$. Therefore for any fixed choice of z there is a maximum value of $\Im(gz)$ as we let g range over G. Let g_0 be an element where we attain that maximum value, i.e. for all $g \in G$ $\Im(gg_0z) \leq Im(g_0z)$. First we let $g = S$, and see that

$$\Im(g_0z) \geq \Im(Sg_0z) = \frac{\Im(g_0(z))}{|g_0z|^2}.$$

We therefore see that $|g_0z|^2 \geq 1$. Let $x_0+iy_0 = g_0z$, so we can see $x_0^2+y_0^2 \geq 1$. It is easy to see for all $n \in \mathbf{Z}$ that $\Im(T^n g_0 z) = \Im(x_0+n+iy_0) = \Im(g_0z)$, but more than that we can choose n such that $|x_0+n| \leq 1/2$. For this choice of n, $T^n g_0 z \in \mathcal{F}$ and of course $T^n g_0 \in G$, which is to say that for any point in $\mathcal{H}$, we can find an element which moves it to $\mathcal{F}$.

Now we pick something in $\mathcal{F}$, say $2i$ and let $\gamma \in \mathrm{SL}_2(\mathbf{Z})$. Therefore $\gamma(2i) \in \mathcal{H}$. Let $g \in G$ such that $g\gamma 2i \in \mathcal{F}$. Therefore $\Im(g\gamma 2i) \geq \sqrt{3}/2$. On the other hand, if $g\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ then we see that we must have $\Im(g\gamma 2i) = \frac{2}{4c^2+d^2} \geq \sqrt{3}/2$. If $c \neq 0$ then $4c^2+d^2 \geq 4$ and so $\Im(g\gamma 2i) \leq 1/2$, so we must have $c=0$ and thus $a = d = \pm 1$. Therefore $g\gamma 2i = 2i+n$ for some $n = \pm b \in \mathbf{Z}$. But since $2i+n \in \mathcal{F}$, we must have $n=b=0$ and thus $\gamma = \pm g^{-1} \in G$. □

The proof above was essentially taken from a blurb of Keith Conrad [2]. There is another proof given, which essentially is the Euclidean algorithm. It should be easy to see that if we give a proof without using $\mathcal{F}$, then we can use that to prove that any $z \in \mathcal{H}$ can be moved to an essentially unique element of $\mathcal{F}$ under the action of $\mathrm{SL}_2(\mathbf{Z})$. That is to say, the fundamental domain for the action of $\mathrm{SL}_2(\mathbf{Z})$ tells us quite a lot about the group $\mathrm{SL}_2(\mathbf{Z})$, or likewise for any discrete subgroup Γ of $\mathrm{SL}_2(\mathbf{R})$ or $\mathrm{PSL}_2(\mathbf{R})$.

Let's return to the topic of modular forms of weight $2k$ for $\mathrm{PSL}_2(\mathbf{Z})$, which are therefore functions f such that $f(z+1) = f(z)$ and $f(-1/z) = z^{2k}f(z)$. A natural way to get the correct transformation factor is the following.

Definition 15. *Define the Eisenstein series of weight* $2k$ *as*

$$G_{2k}(z) = \sum_{\substack{m,n\in\mathbf{Z} \\ (m,n)\neq(0,0)}} \frac{1}{(mz+n)^{2k}}.$$

Theorem 6. *If* $k > 1$, $G_{2k}(z)$ *is a weight* $2k$ *modular form for* $\mathrm{PSL}_2(\mathbf{Z})$.

Proof. The transformation factor is correct because

$$G_{2k}(z+1) = \sum_{\substack{m,n\in\mathbf{Z} \\ (m,n)\neq(0,0)}} \frac{1}{(mz+m+n)^{2k}} = \sum_{\substack{m,n\in\mathbf{Z} \\ (m,n)\neq(0,0)}} \frac{1}{(mz+n)^{2k}} = G_{2k}(z),$$

[2] http://www.math.uconn.edu/~kconrad/blurbs/grouptheory/SL(2,Z).pdf

and

$$G_{2k}(-1/z) = \sum_{\substack{m,n\in\mathbf{Z} \\ (m,n)\neq(0,0)}} \frac{1}{(-m/z+n)^{2k}}.$$

so

$$\frac{1}{z^{2k}} G_{2k}(-1/z) = \sum_{\substack{m,n\in\mathbf{Z} \\ (m,n)\neq(0,0)}} \frac{1}{(nz-m)^{2k}} = \sum_{\substack{m,n\in\mathbf{Z} \\ (m,n)\neq(0,0)}} \frac{1}{(mz+n)^{2k}} = G_{2k}(z).$$

We have however only checked that as a formal series this transformation works. For this to work as a function, we need the series to converge absolutely on $\mathcal{H}$ and uniformly on compact subsets. It is routine to check this is the case for $k > 1$. Once we verify that, it's smooth sailing. We can pass limits through the sums, and therefore derivatives. We have previously verified that linear fractional transformations are holomorphic, and each term in the sum is a linear fractional transformation so the relevant differences in partial derivatives vanish. Moreover, in $\mathcal{F}$, $\Im(z) \to \infty$ if and only if $|z| \to \infty$, so

$$\lim_{\Im(z)\to\infty} G_{2k}(z) = 2\sum_{n\geq 0} \frac{1}{n^{2k}} + \sum_{m\neq 0}\sum_{n\in\mathbf{Z}} \lim_{|z|\to\infty} \frac{1}{(mz+n)^{2k}} = 2\zeta(2k).$$

□

We leave off on this point, but this is only the beginning. As Eichler was rumored to have once said "There are five basic arithmetic operations: addition, subtraction, multiplication, division, and modular forms." There is a lot of interesting mathematics to be mined from Fuchsian groups and their generalizations, and the zeta function suddenly appearing here is only the start.

References

[1] Katok, S. *Fuchsian groups*, University of Chicago Press, Chicago, 1992.

[2] Lovecraft, H. *The call of Cthulhu*, Weird Tales, February 1928.

[3] Riemann, B. *Ueber die Hypothesen, welche der Geometrie zu Grunde liegen*, Abhandlungen der Kniglichen Gesellschaft der Wissenschaften zu Gttingen, **13**, 1867.

[4] Schücker, T. *Forces from Connes' geometry.* Topology and geometry in physics, 285350, Lecture Notes in Phys., 659, Springer, Berlin, 2005.